DIGITAL ACCESS TO SCHOLARSHIP AT HARVARD
DASH.HARVARD.EDU

HARVARD LIBRARY
Office for Scholarly Communication

Whole-exome sequencing and clinical interpretation of FFPE tumor samples to guide precision cancer medicine

Citation
Allen, E. M. V., N. Wagle, P. Stojanov, D. L. Perrin, K. Cibulskis, S. Marlow, J. Jane-Valbuena, et al. 2013. "Whole-exome sequencing and clinical interpretation of FFPE tumor samples to guide precision cancer medicine." Nature medicine 20 (6): 682-688. doi:10.1038/nm.3559. http://dx.doi.org/10.1038/nm.3559.

Published Version
doi:10.1038/nm.3559

Permanent link
http://nrs.harvard.edu/urn-3:HUL.InstRepos:13581165

Terms of Use
This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. Submit a story .

Accessibility

NIH Public Access
Author Manuscript
Nat Med. Author manuscript; available in PMC 2014 December 01.

Published in final edited form as:
Nat Med. 2014 June ; 20(6): 682–688. doi:10.1038/nm.3559.

Whole-exome sequencing and clinical interpretation of FFPE tumor samples to guide precision cancer medicine

Eliezer M. Van Allen[*,1,2], **Nikhil Wagle**[*,1,2], **Petar Stojanov**[2], **Danielle L. Perrin**[2], **Kristian Cibulskis**[2], **Sara Marlow**[1,2], **Judit Jane-Valbuena**[1,2], **Dennis C. Friedrich**[2], **Gregory Kryukov**[2], **Scott L. Carter**[2], **Aaron McKenna**[2,3], **Andrey Sivachenko**[2], **Mara Rosenberg**[2], **Adam Kiezun**[2], **Douglas Voet**[2], **Michael Lawrence**[2], **Lee T. Lichtenstein**[2], **Jeff G. Gentry**[2], **Franklin W. Huang**[1,2], **Jennifer Fostel**[2], **Deborah Farlow**[2], **David Barbie**[1], **Leena Gandhi**[1], **Eric S. Lander**[2], **Stacy W. Gray**[1], **Steven Joffe**[1,4], **Pasi Janne**[1], **Judy Garber**[1], **Laura MacConaill**[1,5], **Neal Lindeman**[1,5], **Barrett Rollins**[1], **Philip Kantoff**[1], **Sheila A. Fisher**[2], **Stacey Gabriel**[**,2], **Gad Getz**[**,#,2,6], and **Levi A. Garraway**[**,#,1,2]

[1]Department of Medical Oncology, Dana-Farber Cancer Institute, Harvard Medical School, 450 Brookline Avenue, Boston, Massachusetts 02115, USA

[2]Broad Institute of MIT and Harvard, 7 Cambridge Center, Cambridge, Massachusetts 02142, USA

[3]Department of Genome Sciences, University of Washington, Seattle, Washington 98195, USA

[4]Children's Hospital Boston, Boston, MA 02115

[5]Brigham and Women's Hospital, Boston, MA 02115

[6]Massachusetts General Hospital Cancer Center and Department of Pathology, Boston, MA 02114

Abstract

Users may view, print, copy, download and text and data- mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use: http://www.nature.com/authors/editorial_policies/license.html#terms

[#]**Corresponding Authors**: Levi A. Garraway, M.D., Ph.D., Department of Medical Oncology, Dana-Farber Cancer Institute, 450 Brookline Avenue, D1542, Boston, MA 02115, USA, Phone: 617-632-6689, Fax: 617- 582-7880, levi_garraway@dfci.harvard.edu, Gad Getz, Ph.D, Cancer Genome Analysis, Broad Institute of MIT and Harvard, 7 Cambridge Center, Cambridge, MA 02142, USA, Phone: 617-714-7471, Fax: 617-800-1152, gadgetz@broadinstitute.org.
[*]These authors contributed equally to this work
[**]These authors contributed equally to this work

Author Contributions

All authors contributed extensively to the work presented in this paper. E.M.V. and N.W. contributed equally to this work. S.G., G.G., and L.A.G. contributed equally to this work. D.L.P., D.C.F., J.F., S.A.F., E.M.V. and S.G. contributed to FFPE sequencing protocols and analysis of sequencing metrics. E.M.V., P.S., K.C., G.K., S.L.C., A.M., A.S., A.K., D.V., M.L., L.T.L., J.G.G., M.R., and G.G. contributed computational analyses for FFPE versus frozen comparisons and analysis of WES data generally. E.M.V., N.W., G.K., F.W.H., S.W.G., S.J., P.J., J.G., L.M., N.L., B.R., P.K., and L.A.G. contributed to clinical analysis and interpretation methods and application. N.W., S.M., J.J., and L.A.G. contributed to experimental follow-up of the JAK3 mutation. D.B. and L.G. contributed clinical input for the patient case. All authors discussed the results and implications and commented on the manuscript at all stages.

Disclosures:

Dr. Garraway and Dr. Wagle are equity holders and consultants in Foundation Medicine. Dr. Garraway is a consultant to Novartis, Millenium/Takeda, and Boehringer Ingelheim, and a recipient of a grant from Novartis. Dr. Getz is a consultant to The Kew Group. Dr. Barbie is a consultant to N-of-1.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Translating whole exome sequencing (WES) for prospective clinical use may impact the care of cancer patients; however, multiple innovations are necessary for clinical implementation. These include: (1) rapid and robust WES from formalin-fixed paraffin embedded (FFPE) tumor tissue, (2) analytical output similar to data from frozen samples, and (3) clinical interpretation of WES data for prospective use. Here, we describe a prospective clinical WES platform for archival FFPE tumor samples. The platform employs computational methods for effective clinical analysis and interpretation of WES data. When applied retrospectively to 511 exomes, the interpretative framework revealed a “long tail” of somatic alterations in clinically important genes. Prospective application of this approach identified clinically relevant alterations in 15/16 patients. In one patient, previously undetected findings guided clinical trial enrollment leading to an objective clinical response. Overall, this methodology may inform the widespread implementation of precision cancer medicine.

Introduction

Massively parallel sequencing approaches such as whole exome sequencing (WES) have elucidated the landscape of genetic alterations in many tumor types and revealed biological insights relevant to clinical contexts[1]. The increased practical availability and decreased cost of tumor genomic profiling has generated opportunities to test the “precision medicine” hypothesis in clinical oncology[2]. In principle, knowledge of alterations in the coding regions of all genes may inform immediate treatment choices and further therapeutic discovery efforts[3].

Most prospective clinical genotyping efforts have utilized “hotspot” genotyping[4–6] or targeted sequencing panels of clinically relevant genes using either fresh frozen or formalin-fixed, paraffin-embedded (FFPE) tissue[7–9]. Pilot studies that apply research-grade massively parallel sequencing technology in focused clinical settings have also been reported[7,10–12], although production-scale efforts have not been demonstrated. Multiple challenges to widespread clinical WES implementation remain. One challenge involves rapidly generating high-quality WES data from archival FFPE tumor material[13]. Another involves clinically interpreting WES data for prospective use that maximizes clinical and biological exploration. A third involves developing a system to interrogate plausibly actionable variants of uncertain significance. Overcoming these challenges should allow rigorous assessment of the value of WES to guide clinical decision-making and inform selected experimental follow-up.

Here we describe an approach to generate high-quality WES data from archival tumor material and validate FFPE WES sequencing data with corresponding frozen WES. We also present a heuristic algorithm that interprets the resulting data for clinical oncologists, and establish the clinical applicability of this interpretation algorithm in a retrospective cohort of 511 cases. Prospective application of this platform in patients with a range of tumor types indicates that this approach enables both biological discovery and clinical trial enrollment. This approach may therefore facilitate widespread application of WES for precision cancer medicine studies.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

Results

Whole exome sequencing of FFPE samples

To produce WES data for clinical use, robust sequencing data must frequently be generated from small quantities of archival FFPE tissue. To test this, DNA was extracted from 99 FFPE samples using the FFPE extraction protocol (Supplementary Table 1, Methods). A comparison of standard WES metrics[14] with 768 non-FFPE samples (394 whole blood, 367 frozen, 7 cell lines) sequenced in parallel demonstrated no significant differences independent of input DNA quantity ($P > 0.05$, Mann-Whitney Test; Fig. 1A–C, Supplementary Table 1). Our lowest successful WES attempts were 13.6ng and 16ng for non-FFPE and FFPE-derived DNA, respectively.

Moreover, improvements in process design (Methods) combined with the "with-bead" approach[14] yielded a time to exome data delivery of 17.4 ± 2.2 days (median ± s.d; $25^{th}/75^{th}$ percentiles 14.3 and 18.6) for FFPE samples received as DNA and 20.1 ± 2.4 days (median ± s.d; $25^{th}/75^{th}$ percentiles 17.5 and 21.2) days for samples received as FFPE tissue blocks (Supplementary Table 2). This turn-around time is compatible with several clinical oncology applications.

We next assessed WES using even smaller amounts of input DNA. Here we achieved > 80% of targeted nucleotides from the hybrid selection reaction, even when only 1ng input DNA was used; equivalent results were seen with FFPE and non FFPE-derived DNA. However, to meet our metrics of ≥ 80% targets ≥ 20x and ≥ 100X mean target coverage, a disproportionate amount of additional sequencing was required due to an increase in the fraction of duplicate molecules in the library.

FFPE and Fresh Frozen samples yield comparable WES results

Next, we sought to compare WES data generated from FFPE and frozen material. We assessed WES data from 11 lung adenocarcinomas for which tumor and adjacent normal tissue were available from matched FFPE (aged ≤ 5 years, Supplementary Table 3, Supplementary Figs 1–2) and frozen samples (Fig. 2A). First, we applied our standard mutation detection pipeline on the tumor-normal pairs (Methods) and considered the concordance of mutation calls observed in FFPE tumors that were observed in frozen tumors, and vice versa. We did not expect identical data given tumor heterogeneity[15] and nucleotide transition artifacts induced by FFPE fixation[16–18]. Moreover, the mean target coverage achieved for the FFPE tumor and adjacent tissue samples were 1.5–2 times that for the corresponding FF samples (Supplementary Fig. 3); as a result, we had increased power to detect mutations in FFPE samples compared to the FF samples[19]. Therefore, we considered the subset of observed exonic mutations in FFPE cases where the depth of coverage afforded sufficient power (> 95%) to detect the mutation in ≥ 2 reads in the matched frozen case, and vice versa. For sufficiently powered sites, 91.5% (2923/3194, 95% confidence interval (CI) ± 0.97) of FFPE mutations validated in patient-matched frozen samples. Similarly, 91.0% (3399/3735, 95% CI ± 0.92) frozen mutations validated in sufficiently powered FFPE samples ($P = 0.47$) (Fig. 2A–C, Supplementary Table 4). Since the mean target coverage in the FFPE cases were higher than their FF counterparts, we then

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

obtained a random subset of reads from each case such that all sites had a maximum coverage of 90X ("downsampling"[19]) and repeated the cross-validation exercise. In this scenario, our FFPE to FF and FF to FFPE validation rates for sufficiently powered sites were 92.6% (2811/3036, 95% CI ± 0.93) and 91.5% (3340/3651, 95% CI ± 0.90), respectively (Supplementary Fig. 4A–B, Supplementary Table 4).

In both FFPE and FF cases from each patient, mutations were observed where there was insufficient power to detect that mutation in the validation cohort after downsampling (Supplementary Fig. 4C, Supplementary Table 4). Demonstrative examples of FFPE mutations that could not be validated in FF counterparts are provided in Supplementary Fig. 5A–C. Overall, these results suggested that the ability to detect base mutations that were sufficiently powered was equivalent regardless of whether frozen or FFPE-derived genomic DNA was used for WES.

We also examined the chromosomal copy number patterns evident in WES data from frozen and FFPE DNA in these 11 cases. In one demonstrative patient, copy ratios for matching exons in FFPE and frozen data correlated ($R^2 = 0.89$, $p < 0.0001$ (Pearson); Fig. 2D–E). This correlation held across all 11 cases, representing 1338859 exons ($R^2 = 0.79$, $p < 0.0001$ (Pearson); Fig. 2F). Thus, WES data obtained from FFPE tumor DNA is comparable to FF WES data, and may equally be used to measure global chromosome copy number information.

Clinical analysis and interpretation of exome sequencing data

Having demonstrated robust WES using FFPE-derived tumor DNA, we next sought to integrate this methodology into a broader framework for clinical interpretation of somatic alterations. We reasoned that a heuristic (rule-based) approach that incorporated prior clinical and scientific knowledge might offer a useful set of organizing principles. By utilizing primary literature, manual curation, and expert opinion, we generated a database of tumor alterations relevant for genomics-driven therapy (TARGET), a database of genes that may have therapeutic, prognostic, and diagnostic implications for cancer patients (Fig. 3B, Supplementary Table 5, Methods). We integrated the resulting 121 TARGET genes with existing open-source resources to create a series of rules that: (i) sort each somatic variant by clinical and biological relevance; (ii) link TARGET genes with additional biologically significant pathways and gene sets; and (iii) demote variants of uncertain significance. Thus, the resulting analytical algorithm used precision heuristics for interpreting the alteration landscape (PHIAL) (Fig. 3A–D, Methods). Beyond annotating variants, PHIAL applies rules that rank variants based on clinical and biological relevance to computationally sort a patient's somatic variants.

The functionality of PHIAL was assessed using 511 patient cases from six prior WES studies[20–25]. Analysis tools (Methods) yielded 258,226 somatic alterations in protein coding genes, of which 135,903 were non-synonymous. Of these, PHIAL identified 1,842 somatic alterations in genes linked to clinical actions (TARGET genes) for 80% (408/511) of the patients (Fig. 3E). Additional descriptive statistics regarding altered genes per patient, stratified by inclusion in databases explored in PHIAL, is available in Supplementary Table 6. PHIAL identified known and highly recurrent actionable findings across this patient

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

cohort. It also revealed a “long tail” of TARGET gene alterations present in small patient subsets that did not reach statistical significance in the individual cohort studies but may have immediate clinical ramifications for individual patients (Fig. 3F). Specifically, 39% (201/511) of the cases had alterations in at least one TARGET gene that was somatically altered in less than 2% of the overall cohort. This finding was reminiscent of similar long tail alteration distributions observed for driver genes in cancer[1].

Since a major near-term goal of precision cancer medicine is to use genetic information to inform clinical trial enrollment, we also systematically queried ClinicalTrials.gov, a centralized registry of publicly and privately supported clinical studies worldwide, for oncology clinical trials linked to TARGET genes. The number of clinical trials including a TARGET gene in the title, the strictest means of identifying clinical trials with a genomic emphasis, grew steadily between 2005 and 2012 (Fig. 3G).

Prospective WES identifies clinically actionable findings across tumor types

To pilot prospective sequencing and clinical interpretation, we performed WES and PHIAL on 16 appropriately consented patients with a range of advanced cancers (Fig. 4A). WES data for 3 of these 16 patients predated the WES protocol described herein, but were included to assess PHIAL output. WES from all patients in the rapid sequencing protocol met our quality control parameters irrespective of tissue processing type (Supplementary Table 7). By completion of the pilot period, sample receipt through data delivery was 16 days.

For these 16 patients, PHIAL revealed 29 unique TARGET genes in the “Investigate Clinical Relevance” category (median: 2, range: 0–5). Although, by definition, alterations in TARGET genes may have implications for clinical decision-making, their actual clinical relevance requires case-by-case evaluation in real time. To facilitate this, every alteration ranked as “Investigate Clinical Relevance” by PHIAL was manually curated to include up-to-date knowledge from databases, literature, and computational algorithms. A standardized, structured annotation was generated for each alteration (Supplementary Note 1), and a level of evidence was assigned to each potential clinical action based on that alteration. These levels of evidence (Table 1) include predictive, prognostic, and diagnostic categories, and encompass validated indications, preclinical evidence, and analytical associations.

Following curation and assignment of levels of evidence, we identified 41 clinically relevant alterations in 15 out of 16 patients. These included standard-of-care findings, such as an *EGFR*^{L858R} mutation in lung adenocarcinoma linked to EGFR inhibitors (predictive for FDA-approved therapies, Level A), and *PIK3CA* alterations that are entry criteria for clinical trials (predictive for therapies in clinical trials, Level A). 46.3% (19/41) of these alterations were based on preclinical evidence for the association of the alteration with response or resistance to FDA-approved therapies or therapies in clinical trials (Level D) (Fig. 4B, Supplementary Table 8).

Multiple unexpected clinically relevant findings were identified in genes not well characterized for the corresponding tumor type. For instance, *CRKL* amplification was observed in a patient with metastatic urothelial carcinoma (Supplementary Fig. 6); this

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

alteration has been predicted to confer resistance to EGFR inhibitors[26] and sensitivity to Src inhibitors[27] in preclinical studies, but had not previously been described in urothelial carcinoma. To accommodate new TARGET genes emerging with future findings, we have made TARGET publically available online and encourage community contributions.

The use of WES in clinical decision-making

The prospective WES framework was used for clinical decision-making in one demonstrative case. A patient with metastatic lung adenocarcinoma underwent standard clinical genetic testing that revealed wild-type *EGFR*, *KRAS* (codon 12 and 13), and *ALK* status. Mass spectrometry testing of 471 alterations in 41 genes[5] revealed an *STK11* frameshift deletion. The patient was started on carboplatin, paclitaxel, and bevacizumab (Fig. 5A). In parallel, we applied the clinical WES platform on the FFPE metastatic tumor sample and germline peripheral blood. PHIAL nominated a *KRAS*^{A146V} mutation, along with alterations in *STK11* (identical to other testing) and *ATM* (Fig. 5A, Supplementary Table 8). *KRAS*^{A146V} is a known activating mutation, though possibly less potent than the codon 12 and 13 mutations[28]. Although activating *KRAS* mutations are found in 15–30% of all patients with non-small cell lung cancer (NSCLC) and commonly in conjunction with *STK11* loss[29], this specific *KRAS* alteration has not been reported in NSCLC[20,30–32]. *KRAS*^{146V} was confirmed using the same FFPE tumor sample in a clinical lab that met Clinical Laboratory Improvement Amendments (CLIA) standards (Knight Diagnostic Laboratories, Oregon), and then returned to the patient's oncologist. After rapidly progressing on combination chemotherapy (Fig. 5B), the patient was enrolled in a phase I CDK4 inhibitor (LY2835219) clinical trial based on preclinical data (Level D) implicating a synthetic lethal relationship between activated *KRAS* and *CDK4*[33]. The patient achieved stable disease (per RECIST 1.1 criteria; 7.9% reduction in tumor volume compared to baseline) and was on therapy for 16 weeks (Fig. 5B–C). Of note, this represented the patient's best and only clinical response to any cancer-directed therapy.

To maximize the potential of clinical WES, we also implemented a procedure to generate experimental evidence for selected Level E (Inferential Association) alterations. An exemplary case involved WES from a patient with metastatic castration-resistant prostate cancer (CRPC) that harbored a *JAK3*^{R870W} missense mutation (Fig. 5D). Activating mutations in *JAK3* have been described in hematological malignancies[34], and *JAK3* inhibitors are available clinically, including the FDA-approved agent tofacitinib. *JAK*^{R870W} has not been previously identified in cancer, and the function of this mutation is unknown.

The crystal structure of JAK3 demonstrates that the arginine at residue 870 directly coordinates the phosphate group of the primary activating tyrosine phosphorylation site (pTyr981)[35] (Fig. 5E). This interaction is expected to pull JAK3 into the active conformation. Indeed, residue 870 is conserved as an arginine or lysine in virtually all JAKs. Given the functional importance of this residue, we hypothesized that this alteration could, in principle, be activating. Thus, this alteration was categorized as Level E (Supplementary Table 8).

We utilized a Ba/F3 system to examine the activity of *JAK3*^{R870W} as compared to *JAK3* WT and a known activating mutation in *JAK3*, A572V[36]. Ba/F3 cells are murine hematopoietic

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

cells dependent of IL-3 for survival. Expression of some oncoproteins substitute for IL-3 signaling, allowing for the growth of Ba/F3 cells in the absence of IL-3. This system has been used extensively to characterize activating mutants of *JAK3* in prior studies[36]. Ba/F3 cells expressing *JAK3*^{R870W} did not achieve IL-3 independent growth following complete IL-3 withdrawal, in contrast to cells expressing a known JAK3 activating mutation (JAK3^{A572V}) or those growing in the presence of IL-3 (Fig. 5F). This suggested *JAK3*^{R870W} is unlikely to be an activating mutation and that *JAK3* inhibitors are unlikely to benefit this patient.

Discussion

This study demonstrates that rapid WES can be applied to FFPE clinical samples, and that robust WES analysis and interpretation can prospectively inform clinical trial enrollment. This approach incorporates new algorithms to identify clinically relevant alterations among numerous somatic events. Furthermore, real-time curation of nominated alterations assigns levels of evidence to the corresponding clinical actions for that alteration in that tumor type. In a proof-of-concept application, we identified at least one clinically relevant alteration in 15 of 16 patients, and showed how such findings can lead to clinical trial enrollment and biological discovery.

Targeted sequencing of clinically relevant gene panels (~100s of genes) have recently become possible from FFPE tumor samples[7], and are increasingly used clinically. However, there are numerous advantages to clinical WES over targeted sequencing. First, as the spectrum of clinically actionable alterations grows[2], targeted sequencing of particular genes are likely to be incomplete: the rapid pace of drug development linked to a growing number of clinically relevant genes will likely outpace the ability to alter targeted sequencing approaches in real time, just as performing clinical WES becomes more facile and cost-efficient. The completeness of clinical WES also enables longitudinal queries if new clinical trials open for previously unrecognized cancer genes not acted on initially.

Furthermore, we expect the volume of inferentially actionable or unknown significance alterations will rise as more patient exomes emerge clinically. Clinical WES allows the generation of deeply annotated genomic data (linked to outcomes and responses) that could be mined to inform TARGET entries. We recognize that the pace of cancer discovery will necessitate continual TARGET updates to ensure its relevance, and we encourage input from the clinical and scientific community to expand and update its content for all to benefit. Methods to aggregate such data in a systems biology approach[37] are being developed to foster functional and clinical follow-up[38,39].

There are ways to improve upon the framework. Efforts to further minimize the input DNA requirement and predict which samples yield successful WES will improve production-level sequencing. This process will be enhanced by pathology review of clinical samples to enrich tumor DNA selection. Improvements in exome-derived copy number algorithms will better distinguish homozygous from heterozygous deletions in stromally admixed tumor samples. Integration of additional profiling technologies (e.g. transcriptome) will provide increasingly complex views of an individual patient's cancer and incorporate other changes (e.g.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

epigenetic) that may have clinical relevance. In parallel, efforts to demonstrate the utility of massively parallel sequencing platforms in larger prospective clinical settings are underway.

PHIAL is heuristic-based; a probabilistic model that assesses alteration clonality with preclinical data may better inform the functional impact of WES findings for individual patients. Even with predictive models, sequencing will frequently identify novel alterations in known genes. Furthermore, relevant information about *known* genomic alterations is constantly changing, and the availability of novel therapies and clinical trials is in rapid flux. Because of this, alteration interpretation presently requires real-time manual curation, which requires dedicated and skilled resources that would benefit from crowdsourcing efforts like we are establishing with TARGET and PHIAL.

Finally, rapid experimental validation of Level E alterations to understand their clinical relevance will require innovations of scale to accelerate functional follow-up. Our experimental efforts described here establish a priority biological evaluation system for one type of functional assessment. A flexible experimental follow-up system to comprehensively assess any alteration will need to be developed.

With the “start to finish” approach for clinical WES described here, it is possible to implement these methods widely and facilitate routine WES in clinical oncology. Once implemented, this will enable the prospective study of patients in trials to determine if large-scale genomic profiling improves patient care and, ultimately, outcomes.

ONLINE METHODS

Patient samples

Tumor and germline sample used for this study were obtained under approved protocols from the Dana-Farber/Harvard Cancer Center Institutional Review Board, the Peter MacCallum Cancer Center Ethics Committee or the Massachusetts Institute of Technology Committee on the Use of Humans as Experimental Subjects. Written informed consent was obtained from all subjects.

Rapid FFPE sequencing

Using industrial best practices in workflow design and a value-add approach, the standard exome workflow was modified to minimize touch points, handoffs, and wasted process steps. Second, optimizations were made to the library construction and in-solution hybridization protocols to enable a 17-hour hybridization reaction, 55 hours shorter than the standard 72-hour hybridization reaction.

FFPE DNA extraction: Paraffin is removed from FFPE sections and cores using CitriSolv™ (Fisher Scientific) followed by ethanol washes, then tissue is lysed overnight at 56°C. Samples are then incubated at 90°C to remove DNA crosslinks, and extraction is performed using Qiagen's QIAamp DNA FFPE Tissue Kit.

Library Construction: This was performed as previously described[14] with the following modifications: initial genomic DNA input into shearing was reduced from 3μg to 10–100ng

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

in 50μL of solution. For adapter ligation, Illumina paired end adapters were replaced with palindromic forked adapters, purchased from Integrated DNA Technologies, with unique 8 base molecular barcode sequences included in the adapter sequence to facilitate downstream pooling. With the exception of the palindromic forked adapters, the reagents used for end repair, A-base addition, adapter ligation, and library enrichment PCR were purchased from KAPA Biosciences in 96-reaction kits. In addition, during the post-enrichment solid phase reversible immobilization (SPRI) bead cleanup, elution volume was reduced to 20μL to maximize library concentration, and a vortexing step was added to maximize the amount of template eluted from the beads. Any libraries with concentrations below 40ng/μl, as measured by a PicoGreen assay automated on an Agilent Bravo, were considered failures and reworked from the start of the protocol.

In-solution hybrid selection: Also performed as previously described[14] with the following modifications to the hybridization reaction: prior to hybridization, any libraries with concentrations >60ng/μL as determined by PicoGreen were normalized to 60ng/μL, and 8.3μL of library was combined with blocking agent, bait, and hybridization buffer. Any libraries with concentrations between 50 and 60ng/μL were normalized to 50ng/μL, and 10.3μL of library was combined with blocking agent, bait, and hybridization buffer. Any libraries with concentrations between 40 and 50ng/μL were normalized to 40ng/μL, and 12.3μL of library was combined with blocking agent, bait, and hybridization buffer. Regardless of library concentration range, the same volume of blocking agent and bait previously described[14] were used, and hybridization buffer volume was adjusted to equal the combined volume of library, blocking agent, and bait. Finally, the hybridization reaction was reduced to 17 hours with no changes to the downstream capture protocol.

Preparation of libraries for cluster amplification and sequencing: After post-capture enrichment, libraries were quantified using PicoGreen (automated assay on the Agilent Bravo), normalized to equal concentration on the Perkin Elmer MiniJanus, and pooled by equal volume on the Agilent Bravo. Library pools were then quantified using quantitative PCR (kit purchased from KAPA Biosystems) with probes specific to the ends of the adapters; this assay was automated using Agilent's Bravo liquid handling platform. Based on qPCR quantification, libraries were normalized to 2nM, then denatured using 0.2 N NaOH on the Perkin-Elmer MiniJanus. After denaturation, libraries were diluted to 20pM using hybridization buffer purchased from Illumina.

Cluster amplification and sequencing: Cluster amplification of denatured templates was performed according to the manufacturer's protocol (Illumina) HiSeq v3 cluster chemistry and flowcells, as well as Illumina's Multiplexing Sequencing Primer Kit. Flowcells were sequenced using HiSeq 2000 v3 Sequencing-by-Synthesis Kits, then analyzed using RTA v. 1.12.4.2 or later. Each pool of whole exome libraries was run on paired 76bp runs, and 8 base index sequencing read was performed to read molecular indices, across the number of lanes needed to meet coverage for all libraries in the pool.

Statistical analysis of raw sequencing metrics: All analyses of raw sequencing metrics were performed using the R statistical package. Sample size was established by incorporating all available FFPE samples sequenced under the FFPE sequencing protocol by the time of

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

analysis freeze (n = 99). Significance between two means (FFPE and non-FFPE samples for the sequencing metrics) was calculated with the two-tailed Mann-Whitney test given the non-normal distribution of values. $P < 0.05$ was considered significant.

Analysis and Interpretation

DNA Assembly and Quality Control—Sequence data processing: Exomes sequence data processing was performed using established pipelines at the Broad Institute. A BAM file was produced with the Picard pipeline (http://picard.sourceforge.net/), which aligns the tumor and normal sequences to the hg19 human genome build using Illumina sequencing reads. The BAM was uploaded into the Firehose pipeline (http://www.broadinstitute.org/cancer/cga/Firehose), which manages input and output files to be executed by GenePattern[40]. Whole exome sequencing BAM files for data from this study cases will be deposited in dbGAP (phs000488 for lung adenocarcinoma cases; phs number pending for clinical cases).

Sequencing quality control: Quality control modules within Firehose were applied to all sequencing data for comparison of the origin for tumor and normal genotypes and confirm fingerprinting concordance. Cross-contamination of samples was estimated using ContEst[41], to confirm that neither tumor nor germline sample had > 3% contamination. SNP fingerprints from each lane of a tumor/normal pair were cross-checked to confirm concordance, and non-matching lanes were removed from analysis.

Somatic alteration identification and annotation—The MuTect algorithm[19] was applied to identify somatic single-nucleotide variants in targeted exons. Indelocator (http://www.broadinstitute.org/cancer/cga/indelocator) was applied to identify small insertions or deletions. Annotation of identified variants was done using Oncotator (http://www.broadinstitute.org/cancer/cga/oncotator). Rearrangments were identified using dRanger (http://www.broadinstitute.org/cancer/cga/dranger). Copy ratios were calculated for each hybrid capture bait by dividing the tumor coverage by the median coverage obtained in a set of reference normal samples[42]. The resulting copy ratios were segmented using the circular binary segmentation algorithm[43]. Genes in copy ratio regions with segment means of greater than $\log_2(4)$ were evaluated for focal amplifications given the potential clinical significance of a large focal event. Genes in regions with segment means of less than $\log_2(0.5)$ were evaluated for hemizygous or homozygous deletions, since either broad or focal deletions may involve genes with clinical relevance. RefSeq[44] was used to identify the genes that reside in the chromosomal coordinates demarcated by the segment start and end points.

Cross-validation of FFPE and Fresh Frozen mutation data—FFPE sections were received as 15μ slices (9 per sample), aged 2007–2009. All FFPE samples were sequenced as described above with 100ng of input DNA. Frozen tumor samples were sequenced according to established methods[14]. All downstream computational analysis methods for assembly, alignment, mutation, and copy number alteration identification were identical to the pipelines described above. For the downsampling experiment, MuTect was re-run on all the cases with the “downsample_to_coverage” parameter set to 90. Mutations in intronic

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

regions were excluded. For cross-validation of mutations, validation power was defined as the probability to observe at least two alternative allele reads in the validation sample (given the allelic fraction, coverage in validation sample at that site, and the assumption that the mutation should be present there).

Statistical analysis of FFPE and frozen tissue: Two-tailed Fisher's exact test was used to test the statistical significance of the contingency table represented by tissue type (FFPE or frozen) and validation status. Pearson correlation was performed on $\log_2$(Target Copy Ratio) segment mean data for FFPE and frozen exon targets and significance was calculated using Pearson's product moment correlation coefficient. Sample size for exons from all 11 cases (n = 1,338,859) greatly exceeded the minimum sample size needed to determine a linear correlation coefficient of 0.8 with power of 0.8 and significance level of 0.05. The variance estimate among FFPE (0.054) and frozen (0.049) copy number signal data was similar. Whole exome sequencing data for lung adenocarcinoma cases will be deposited in dbGAP (phs000488).

Clinical Gene Database (TARGET)—The TARGET (tumor alterations relevant for genomics-driven therapy) database included genes that, when altered somatically in cancers, met one of three criteria:

1. Alterations in the gene predicted resistance and/or sensitivity to specific therapies
2. Alterations in the gene had prognostic significance in a cancer type
3. Alterations in the gene had diagnostic significance in a cancer type.

To build this database, we performed a systematic review of the primary literature, manually curated specific genes based on clinician input, and consulted expert opinion. This resulted in a list of 121 genes that met at least one of the three criteria required for entry into the TARGET database (Supplementary Table 5, available for download at www.broadinstitute.org/cancer/cga/target).

Somatic heuristic algorithm for interpretation (PHIAL)—Each somatic variant was scored individually using a series of rules, and then was considered in aggregate to determine relationships between alterations in the same patient (e.g. linked pathways). First, variants in TARGET are ranked highest, with scoring modifications for known mutational hotspots (e.g. *BRAF* V600E), missense mutations in protein kinase regions, and copy number alterations with directionality known to have clinical impact (e.g. *PTEN* deletion). To assign maximum granularity between alterations, additional rules assign priority based on presence of recurrent alterations in the Cancer Gene Census[45], presence in the pathway of concurrently altered actionable genes in the same sample using curated cancer pathways from MSigDB[46], presence in known cancer pathways, gene sets, or modules identified by MSigDB, and finally presence in COSMIC[30]. All code for PHIAL was implemented using the R statistical package language and is available online (www.broadinstitute.org/cancer/cga/phial).

Visual Representation—A decision support tool built around the results was developed to allow curation team members and clinicians to engage the data with web-based resources

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

integrated directly into the patient's results. The tool is built to convey effective clinical review with the minimum manual steps so that such a process can be scaled rapidly. The report structure was implemented using the Nozzle R package[47]. All clinically actionable relevant somatic variants were linked to search criteria in ClinicalTrials.gov.

Curation—Somatic alterations nominated by PHIAL as "investigate clinical relevance" were assigned for curation by a team of oncology and genomics experts charged with answering a series of structured questions pertaining to each nominated variant to facilitate final review (Supplementary Fig. 6). A curated alteration required review of published data to determine which level of evidence could be assigned to a clinical action for the alteration (Table 1).

Clinical Trial Data Analyses—Clinicaltrials.gov (http://clinicaltrials.gov/) was accessed on 2/19/2013 and the search entry "cancer" was used to extract all cancer-related clinical trials in the database. Duplicated trial entries and trials designated as "Terminated" or "Withdrawn" were excluded. Provided trial start dates date (by year) were used to select all trials that were initiated between 2005 and 2012, and trial titles were queried using string matching in R for those that specifically mention TARGET genes in the title of the trial.

Ba/F3 Experimental Methods

Cell Culture—HEK293T cells were maintained in DMEM (Gibco) with 10% (vol/vol) fetal bovine serum (Gibco). Ba/F3 cells (a kind gift of Dr. Andrew Lane at Dana-Farber Cancer Institute) were maintained in RPMI 1640 (Gibco) with 10% FBS and 10 ng/mL mouse interleukin-3 (IL-3; Prospec).

Retroviral infections—The wild-type JAK3 cDNA cloned in the pDONR223 vector was obtained from The Broad Institute RNAi Consortium. JAK3 mutations were generated by site-directed mutagenesis using the QuikChange Lightning Mutagenesis Kit (Stratagene), and verified by full sequencing of the JAK3 cDNA insert. WT and mutant cDNAs were recombined into a Gateway adapted MSCV-puromycin vector (kind gift of Dr. Akinori Yoda at Dana Farber Cancer Institute) using the Gateway LR Clonase kit. Ecotropic viruses were produced by cotransfection of MSCV-constructs with pCL-Eco vector (Imgenex) in 293T cells. Ba/F3 cells were plated in 6-well plates at a 30% confluency, and spin infected at 800 × g for 90 min at 33 °C in the presence of 8 μg/mL polybrene (hexadimethrine bromide; Sigma). The same infection protocol was repeated 24 h later. Upon completion, the viral supernatant was removed and fresh media added. Twenty-four hours after the media change, Ba/F3 cells were subjected to a 3-day puromycin selection (2 μg/mL) in the presence of IL-3. Expression of ectopic JAK3 protein was verified by immunoblot analysis using a primary antibody against phospho-JAK3 (Cell Signaling #5031).

IL-3 Depletion—Ba/F3 cells and Ba/F3 cells expressing WT and mutant forms of JAK3 were seeded in 25 cm^2 vented-cap flasks at 20,000 cells/ml in a total volume of 5mls in the absence of IL-3 to select IL-3–independent cells. Cells were grown in the absence of IL-3 over several weeks. In parallel, Ba/F3 cells were maintained in 10ng/ml IL-3 throughout as a

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

positive control. Cell counts were recorded every 4 days using ViCell counter and split as needed.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank the patients and clinicians for their participation in this project. We thank the Broad Genomics Platform (specifically K. Anderka, A. Cheney, E. Wheeler, T. Mason, and C. Crawford). Finally, we thank A. Lane and A. Yoda for their contributions to the *JAK3* experimental work.

Financial Support:

This work was supported by Starr Cancer Foundation (L.A.G.), Prostate Cancer Foundation (L.A.G.), NHGRI CSER grant (L.A.G.), National Cancer Institute (L.A.G.), Department of Defense (L.A.G.), NIH U24CA143845 grant (G.G.) and the Dana-Farber Leadership Council (E.M.V.)

References

1. Garraway LA, Lander ES. Lessons from the cancer genome. Cell. 2013; 153:17–37. [PubMed: 23540688]
2. Garraway LA. Genomics-Driven Oncology: Framework for an Emerging Paradigm. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2013
3. Garraway LA, Janne PA. Circumventing cancer drug resistance in the era of personalized medicine. Cancer discovery. 2012; 2:214–226. [PubMed: 22585993]
4. Thomas RK, Baker AC, Debiasi RM, Winckler W, Laframboise T, et al. High-throughput oncogene mutation profiling in human cancer. Nature genetics. 2007; 39:347–351. [PubMed: 17293865]
5. MacConaill LE, Campbell CD, Kehoe SM, Bass AJ, Hatton C, et al. Profiling critical cancer gene mutations in clinical tumor samples. PloS one. 2009; 4:e7887. [PubMed: 19924296]
6. Dias-Santagata D, Akhavanfard S, David SS, Vernovsky K, Kuhlmann G, et al. Rapid targeted mutational analysis of human tumours: a clinical platform to guide personalized cancer medicine. EMBO molecular medicine. 2010; 2:146–158. [PubMed: 20432502]
7. Wagle N, Berger MF, Davis MJ, Blumenstiel B, Defelice M, et al. High-throughput detection of actionable genomic alterations in clinical tumor samples by targeted, massively parallel sequencing. Cancer discovery. 2012; 2:82–93. [PubMed: 22585170]
8. Lipson D, Capelletti M, Yelensky R, Otto G, Parker A, et al. Identification of new ALK and RET gene fusions from colorectal and lung cancer biopsies. Nature medicine. 2012; 18:382–384.
9. Beltran H, Yelensky R, Frampton GM, Park K, Downing SR, et al. Targeted next-generation sequencing of advanced prostate cancer identifies potential therapeutic targets and disease heterogeneity. European urology. 2013; 63:920–926. [PubMed: 22981675]
10. Roychowdhury S, Iyer MK, Robinson DR, Lonigro RJ, Wu YM, et al. Personalized oncology through integrative high-throughput sequencing: a pilot study. Science translational medicine. 2011; 3 111ra121.
11. Craig DW, O'Shaughnessy JA, Kiefer JA, Aldrich J, Sinari S, et al. Genome and transcriptome sequencing in prospective metastatic triple-negative breast cancer uncovers therapeutic vulnerabilities. Molecular cancer therapeutics. 2013; 12:104–116. [PubMed: 23171949]
12. Kerick M, Isau M, Timmermann B, Sultmann H, Herwig R, et al. Targeted high throughput sequencing in clinical cancer settings: formaldehyde fixed-paraffin embedded (FFPE) tumor tissues, input amount and tumor heterogeneity. BMC medical genomics. 2011; 4:68. [PubMed: 21958464]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

13. Goetz L, Bethel K, Topol EJ. Rebooting cancer tissue handling in the sequencing era: toward routine use of frozen tumor tissue. JAMA : the journal of the American Medical Association. 2013; 309:37–38. [PubMed: 23280221]
14. Fisher S, Barry A, Abreu J, Minie B, Nolan J, et al. A scalable, fully automated process for construction of sequence-ready human exome targeted capture libraries. Genome biology. 2011; 12:R1. [PubMed: 21205303]
15. Gerlinger M, Rowan AJ, Horswell S, Larkin J, Endesfelder D, et al. Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. The New England journal of medicine. 2012; 366:883–892. [PubMed: 22397650]
16. Pao W, Ladanyi M, Miller VA. Erlotinib in lung cancer. The New England journal of medicine. 2005; 353:1739–1741. author reply 1739–1741. [PubMed: 16240471]
17. Williams C, Ponten F, Moberg C, Soderkvist P, Uhlen M, et al. A high frequency of sequence alterations is due to formalin fixation of archival specimens. The American journal of pathology. 1999; 155:1467–1471. [PubMed: 10550302]
18. Spencer DH, Sehn JK, Abel HJ, Watson MA, Pfeifer JD, et al. Comparison of Clinical Targeted Next-Generation Sequence Data from Formalin-Fixed and Fresh-Frozen Tissue Specimens. The Journal of molecular diagnostics : JMD. 2013
19. Cibulskis K, Lawrence MS, Carter SL, Sivachenko A, Jaffe D, et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. Nature biotechnology. 2013
20. Imielinski M, Berger AH, Hammerman PS, Hernandez B, Pugh TJ, et al. Mapping the hallmarks of lung adenocarcinoma with massively parallel sequencing. Cell. 2012; 150:1107–1120. [PubMed: 22980975]
21. Banerji S, Cibulskis K, Rangel-Escareno C, Brown KK, Carter SL, et al. Sequence analysis of mutations and translocations across breast cancer subtypes. Nature. 2012; 486:405–409. [PubMed: 22722202]
22. Hodis E, Watson IR, Kryukov GV, Arold ST, Imielinski M, et al. A landscape of driver mutations in melanoma. Cell. 2012; 150:251–263. [PubMed: 22817889]
23. Barbieri CE, Baca SC, Lawrence MS, Demichelis F, Blattner M, et al. Exome sequencing identifies recurrent SPOP, FOXA1 and MED12 mutations in prostate cancer. Nature genetics. 2012; 44:685–689. [PubMed: 22610119]
24. Stransky N, Egloff AM, Tward AD, Kostic AD, Cibulskis K, et al. The mutational landscape of head and neck squamous cell carcinoma. Science. 2011; 333:1157–1160. [PubMed: 21798893]
25. Lohr JG, Stojanov P, Lawrence MS, Auclair D, Chapuy B, et al. Discovery and prioritization of somatic mutations in diffuse large B-cell lymphoma (DLBCL) by whole-exome sequencing. Proceedings of the National Academy of Sciences of the United States of America. 2012; 109:3879–3884. [PubMed: 22343534]
26. Cheung HW, Du J, Boehm JS, He F, Weir BA, et al. Amplification of CRKL induces transformation and epidermal growth factor receptor inhibitor resistance in human non-small cell lung cancers. Cancer discovery. 2011; 1:608–625. [PubMed: 22586683]
27. Natsume H, Shinmura K, Tao H, Igarashi H, Suzuki M, et al. The CRKL gene encoding an adaptor protein is amplified, overexpressed, and a possible therapeutic target in gastric cancer. Journal of translational medicine. 2012; 10:97. [PubMed: 22591714]
28. Janakiraman M, Vakiani E, Zeng Z, Pratilas CA, Taylor BS, et al. Genomic and biological characterization of exon 4 KRAS mutations in human cancer. Cancer research. 2010; 70:5901–5911. [PubMed: 20570890]
29. Carretero J, Shimamura T, Rikova K, Jackson AL, Wilkerson MD, et al. Integrative genomic and proteomic analyses identify targets for Lkb1-deficient metastatic lung tumors. Cancer cell. 2010; 17:547–559. [PubMed: 20541700]
30. Forbes SA, Bindal N, Bamford S, Cole C, Kok CY, et al. COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. Nucleic acids research. 2011; 39:D945–950. [PubMed: 20952405]
31. Ding L, Getz G, Wheeler DA, Mardis ER, McLellan MD, et al. Somatic mutations affect key pathways in lung adenocarcinoma. Nature. 2008; 455:1069–1075. [PubMed: 18948947]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

32. Govindan R, Ding L, Griffith M, Subramanian J, Dees ND, et al. Genomic landscape of non-small cell lung cancer in smokers and never-smokers. Cell. 2012; 150:1121–1134. [PubMed: 22980976]

33. Puyol M, Martin A, Dubus P, Mulero F, Pizcueta P, et al. A synthetic lethal interaction between K-Ras oncogenes and Cdk4 unveils a therapeutic strategy for non-small cell lung carcinoma. Cancer cell. 2010; 18:63–73. [PubMed: 20609353]

34. Levine RL. JAK-mutant myeloproliferative neoplasms. Current topics in microbiology and immunology. 2012; 355:119–133. [PubMed: 21823028]

35. Boggon TJ, Li Y, Manley PW, Eck MJ. Crystal structure of the Jak3 kinase domain in complex with a staurosporine analog. Blood. 2005; 106:996–1002. [PubMed: 15831699]

36. Malinge S, Ragu C, Della-Valle V, Pisani D, Constantinescu SN, et al. Activating mutations in human acute megakaryoblastic leukemia. Blood. 2008; 112:4220–4226. [PubMed: 18755984]

37. Gonzalez-Angulo AM, Hennessy BT, Mills GB. Future of personalized medicine in oncology: a systems biology approach. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2010; 28:2777–2783. [PubMed: 20406928]

38. Yeh P, Chen H, Andrews J, Naser R, Pao W, et al. DNA-mutation Inventory to Refine and Enhance Cancer Treatment (DIRECT): A catalogue of clinically relevant cancer mutations to enable genome-directed cancer therapy. Clinical cancer research : an official journal of the American Association for Cancer Research. 2013

39. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. Cancer discovery. 2012; 2:401–404. [PubMed: 22588877]

Online Methods References

40. Reich M, Liefeld T, Gould J, Lerner J, Tamayo P, et al. GenePattern 2.0. Nature genetics. 2006; 38:500–501. [PubMed: 16642009]

41. Cibulskis K, McKenna A, Fennell T, Banks E, DePristo M, et al. ContEst: estimating cross-contamination of human samples in next-generation sequencing data. Bioinformatics. 2011; 27:2601–2602. [PubMed: 21803805]

42. Chiang DY, Getz G, Jaffe DB, O'Kelly MJ, Zhao X, et al. High-resolution mapping of copy-number alterations with massively parallel sequencing. Nature methods. 2009; 6:99–103. [PubMed: 19043412]

43. Olshen AB, Venkatraman ES, Lucito R, Wigler M. Circular binary segmentation for the analysis of array-based DNA copy number data. Biostatistics. 2004; 5:557–572. [PubMed: 15475419]

44. Pruitt KD, Tatusova T, Klimke W, Maglott DR. NCBI Reference Sequences: current status, policy and new initiatives. Nucleic acids research. 2009; 37:D32–36. [PubMed: 18927115]

45. Futreal PA, Coin L, Marshall M, Down T, Hubbard T, et al. A census of human cancer genes Nature reviews. Cancer. 2004; 4:177–183. [PubMed: 14993899]

46. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proceedings of the National Academy of Sciences of the United States of America. 2005; 102:15545–15550. [PubMed: 16199517]

47. Gehlenborg N, Noble MS, Getz G, Chin L, Park PJ. Nozzle: a report generation toolkit for data analysis pipelines. Bioinformatics. 2013

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

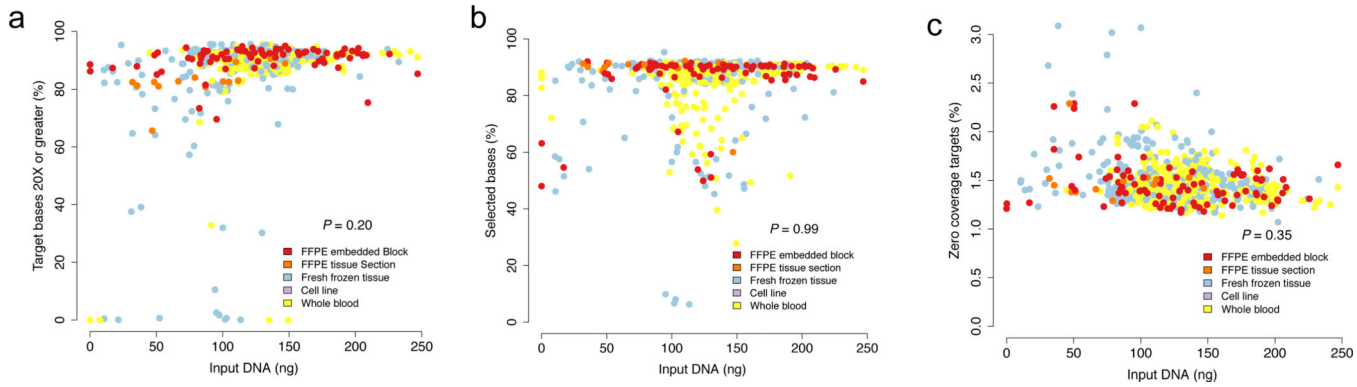

Figure 1. FFPE and Frozen sequencing metrics

The percentage of target bases covered at 20X, percent selected bases, and percent of zero coverage targets in FFPE (n = 99) and non-FFPE tissue (n = 768) (1A–C). Additional quality control metrics for all 867 cases are available in Supplementary Table 1. No statistically significant difference between FFPE and non-FFPE tissue is observed in these three metrics ($P > 0.05$; two-sided Mann-Whitney test).

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

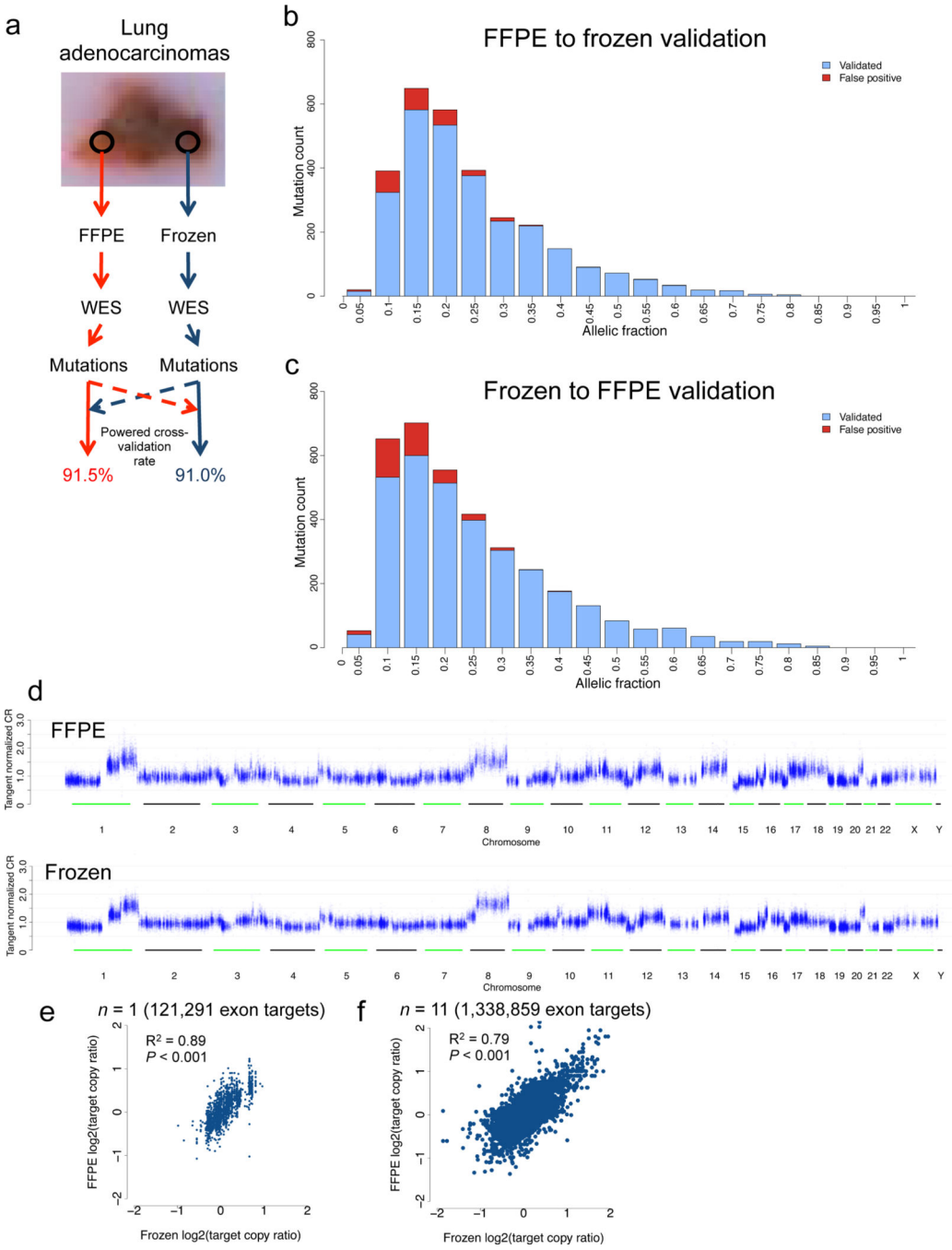

Figure 2. FFPE and frozen data yield comparable alteration data

FFPE and frozen tissue were extracted from identical tumor samples and analyzed for cross-validation of mutations where there was sufficient power to detect the mutation in the validation sample (A). FFPE to frozen and frozen to FFPE validation rates binned by allelic fractions demonstrate similar validation and false positive rates between the two groups (B–C). Copy number profiles derived from exomes of the same tumor in either FFPE or frozen tissue yield comparable results (R^2 (Pearson) = 0.89; $P < 0.001$) (D–E). When comparing

the FFPE and frozen segment means for all exons across 11 patients, the R^2 (Pearson) = 0.79 ($P < 0.001$) (F).

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

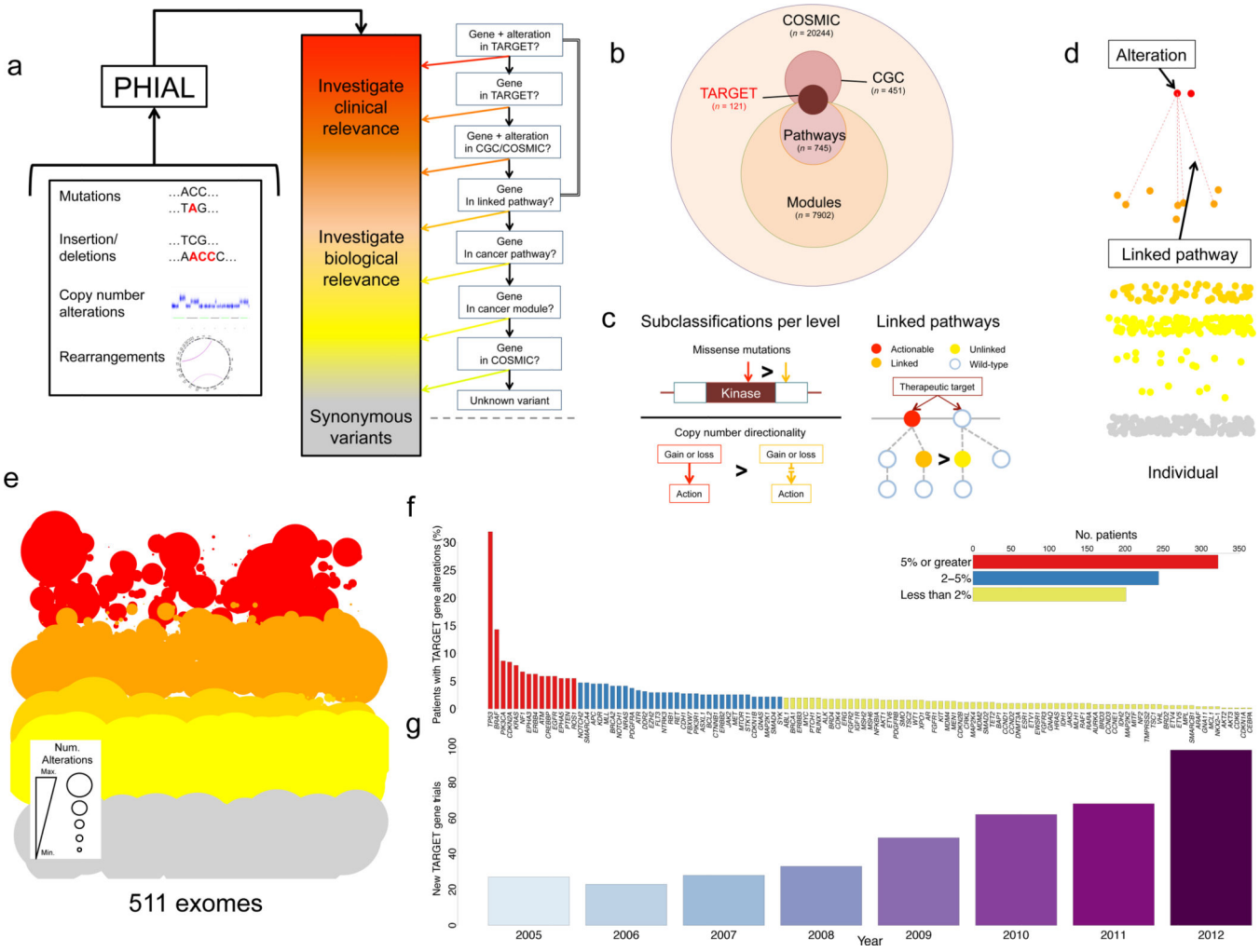

Figure 3. PHIAL reveals the "long tail" of clinically relevant events

PHIAL takes as input somatic alterations and uses heuristics to assign clinical and biological significance to each alteration (A). PHIAL uses the TARGET database, a curated set of genes that are linked to predictive, prognostic, and/or diagnostic clinical actions when somatically altered in cancers. (B). PHIAL utilizes additional rules to maximize exome data for individuals, including knowledge about kinase domains, copy number directionality, and two-hit pathway events (C). The resulting data is visualized for individual or cohort-level information with this demonstrative PHIAL "gel". Each alteration is a point sorted by PHIAL score (top are of highest clinical relevance), color coded by potential clinical relevance (red), biological relevance (orange), pathway relevance (yellow), or synonymous variants (gray) (D). A PHIAL "gel" for 511 patient exomes spanning six different disease types (n = 258,226 total somatic alterations). The size of the point is proportional to the number of times a given gene arises at that PHIAL score level. (E). This approach highlights the "long tail" of potentially clinically relevant alterations in TARGET genes (n = 121) that may be present in an individual patient but does not occur sufficiently to be labeled a biological driver across a cohort. The majority of events occur in genes that individually are altered in less than 2% of the overall cohort (F). New cancer clinical trials with TARGET

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

genes specifically integrated into the study per ClinicalTrials.gov over a seven-year period (G).

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

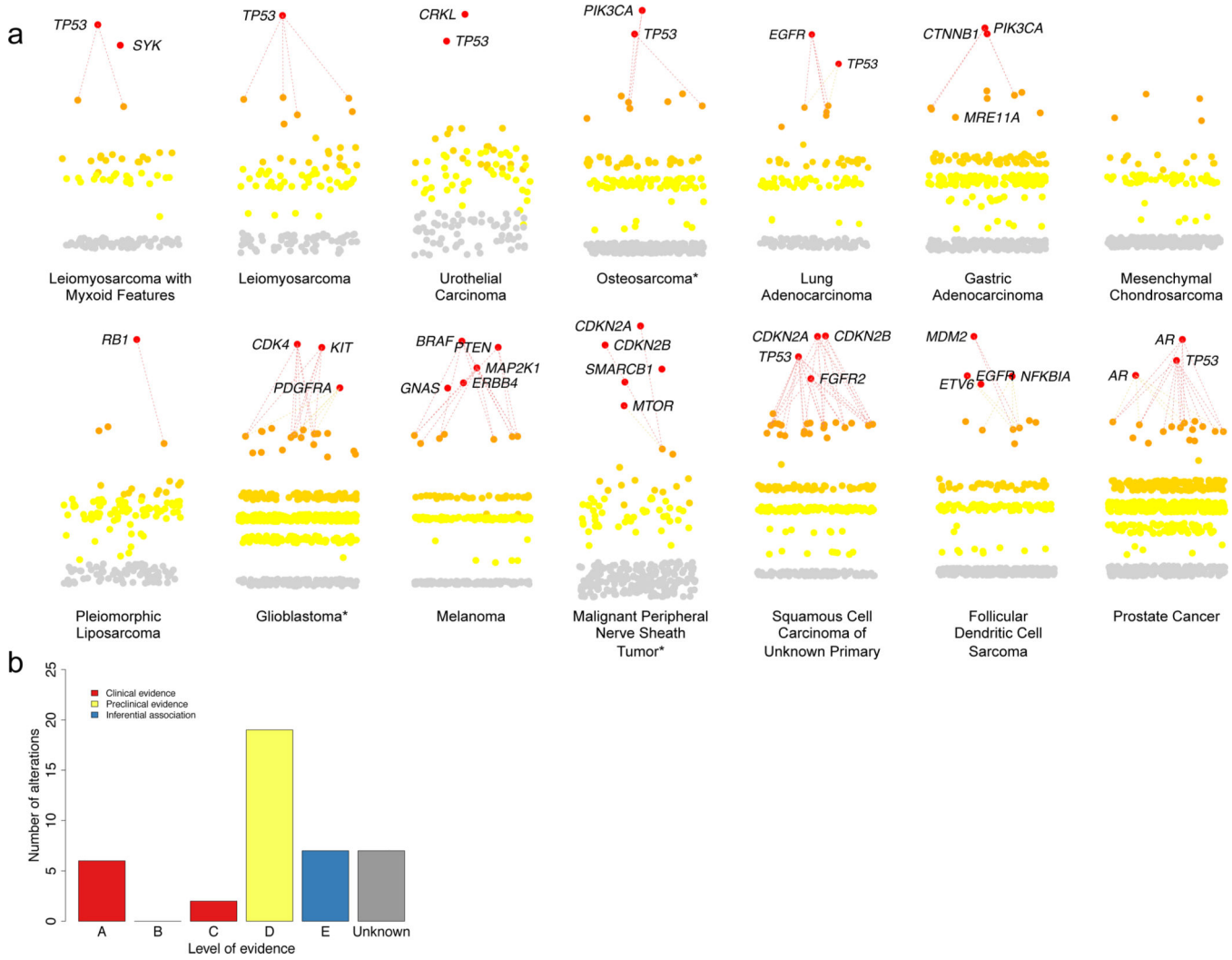

Figure 4. Clinically relevant findings from individual patients

PHIAL results for 14 patients with a spectrum of malignancies, highlighting nominated clinically actionable alterations in 13 of 14 patients (A). Using the level of evidence schematic (Table 1), all nominated alterations for patients in this study were manually curated and assigned a level of evidence (B, Supplementary Table 7). *Denotes patient sequencing data that predated the rapid WES protocol.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

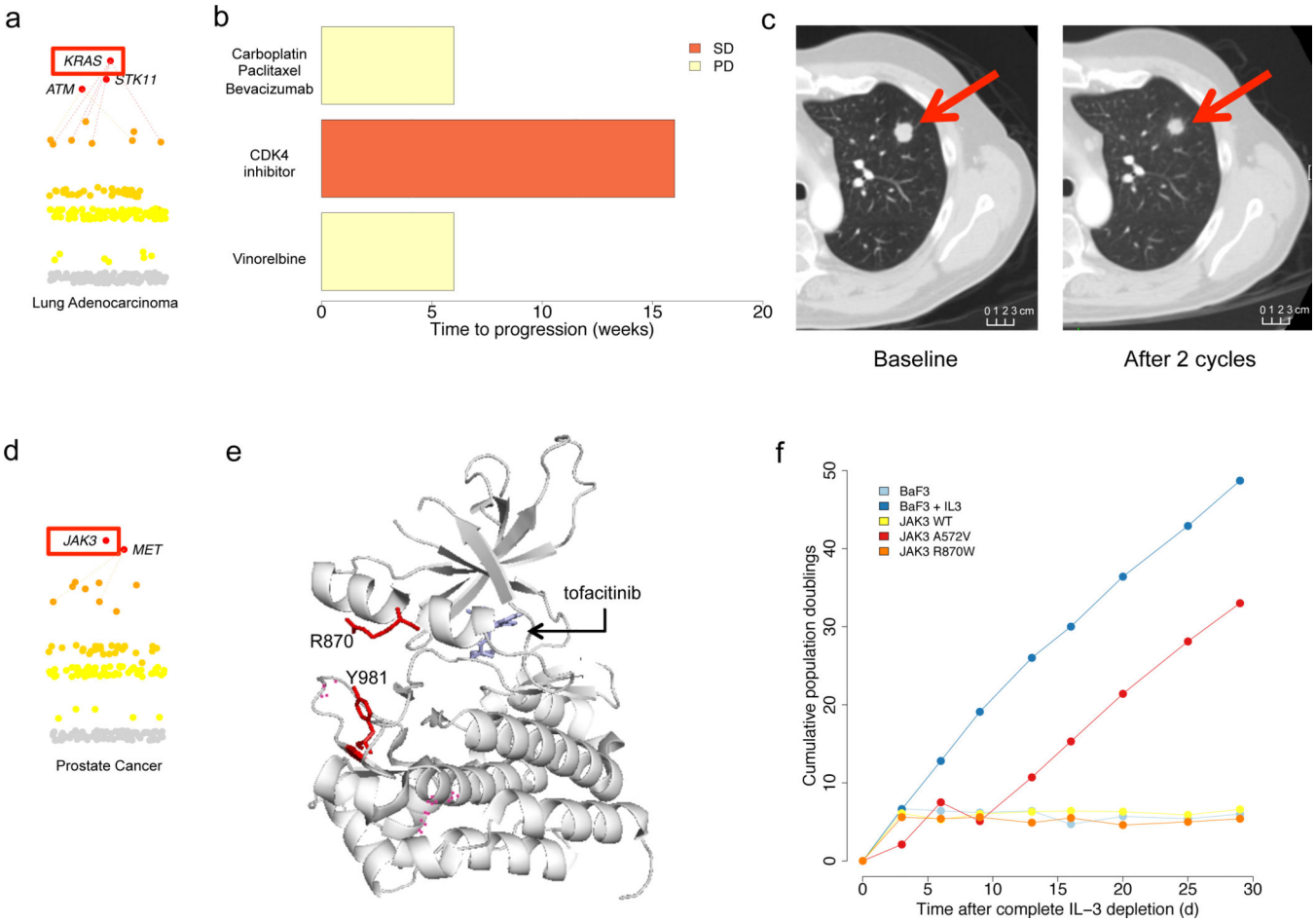

Figure 5. Clinical sequencing informs clinical trial enrollment and experimental discovery

The PHIAL output and treatment course for a patient with metastatic lung adenocarcinoma is shown, with the integration of clinical WES occurring during the patient's first-line therapy allowing subsequent clinical trial enrollment (A). The patient's time to relapse data for the three treatment regimens received demonstrate that the best and only clinical response occurred with the CDK4 inhibitor (B). Radiographic imaging demonstrates a small reduction in a representative metastatic focus for the patient on the CDK4 inhibitor trial after two cycles of therapy consistent with stable disease (cm: centimeter; measurement is 1.7 × 1.5 cm for baseline mass and 1.3 × 1.3 cm for two month interval scan of the same mass). Per RECIST criteria, overall tumor reduction was 7.9% (C). For another patient, PHIAL nominated a *JAK3* missense mutation (D), and given its location in the kinase domain near alterations previously defined as activating, was considered to have inferential evidence (Level E) for being clinically actionable. The crystal structure of JAK3 demonstrates that the arginine at residue 870 directly coordinates the phosphate group of the primary activating tyrosine phosphorylation site (E). To better characterize this alteration, experimental follow-up of this alteration was performed in a Ba/F3 system. Overexpression of the patient's *JAK3* mutation did not suggest an activating phenotype or further consideration of JAK3 inhibitor clinical trial enrollment (F).

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

Table 1

Levels of Evidence

To assign the strictest possible definition of evidence linking an alteration to a clinical action for each given variant, all highest scoring variants are curated with the aim of assigning each a level (A through E) depending on the category of action. Inferential associations are based on homology, computational data, structural information, or pathway involvement. Those that do not meet criteria for any level of evidence are considered unknown variants.

CATEGORY:	LEVEL A	LEVEL B	LEVEL C	LEVEL D	LEVEL E
Predictive – FDA-approved therapies	There is a **validated association** between this alteration and response/resistance to this agent ***for this indication***	There is **limited clinical evidence** (early or conflicting data) for an association between this alteration and response/resistance to this agent ***in this tumor type***	There is **clinical evidence** for an association between this alteration and response/resistance to this agent in ***another tumor type ONLY***	There is **preclinical evidence** for an association between this alteration and response/resistance to this agent	There is an **inferential association** between this alteration and response/resistance to this agent
Predictive – Therapies in clinical trials	This alteration is used or has been used as an **eligibility criterion** for clinical trials of this agent or class of agents	There is **limited clinical evidence** (early or conflicting data) for an association between this alteration and response/resistance to this agent or class of agents ***in this tumor type***	There is **clinical evidence** for an association between this alteration and response/resistance to this agent or class of agents in ***another tumor type ONLY***	There is **preclinical evidence** for an association between this alteration and response/resistance to this agent or class of agents	There is a **inferential association** between this alteration and response/resistance to this agent or class of agents
Prognostic	There is a **validated association** between this alteration and prognosis in this tumor type	There is **limited evidence** for an association between this alteration and prognosis in this tumor type			
Diagnostic	There is a **validated association** between this alteration and a diagnosis	There is **limited evidence** for an association between this alteration and a diagnosis			

Nat Med. Author manuscript; available in PMC 2014 December 01.